A Novel design of a Quadruple Floating-point Fused Multiply-Add Unit

[1]PROF.A.S.DESHMUKH, [2]PROF.M.A.DESHMUKH,[3]PROF.D.P.DONGRE
[1]Assistant Professor, [2]Assistant Professor , [3]Assistant Professor
Deptt. Of Electronic & Telecommunication Engineering
Govindrao Wanjari College of Engg. & Technology, Nagpur, Maharashtra, India

Abstract : Floating-point unit (FPU) is one of the most important custom applications needed in most hardware designs. A Quadruple Precision (QP) floating-point arithmetic is one of the important trend, which helps to increase the precision of floating-point arithmetic and improve the performance of some scientific applications. According to the characteristic of quadruple precision (QP) floating-point data format, a high performance QPFMA is designed and realized, which supports multiple floating-point arithmetic. High precision and high performance floating-point unit is an important research object of high performance microprocessor design. QPFMA design, satisfying the requirements of high performance processor design.

I. INTRODUCTION

Quadruple Precision Floating-point fused multiply add unit (FMA) is one of the most important blocks useful in many computations which involve the accumulation of products such as scientific and engineering applications. Many algorithms are developed on floating-point fused multiply add unit to decrease its latency also important as it increases its accuracy and performance [2], [10], [18], [25], [31].

The IEEE 754-2008 floating-point standard [1]has included quadruple precision (QP) floating-point data type (binary128) to support high precision floating-point computation. This helps to improve the precision and reoccurrence of floating-point results and enhance the stability of numerical algorithms. And also the floating-point fused multiply-add (FMA) operation is taken as one of the basic operations. So it has become one of the hot topics on how to design and implement high precision and high performance floating point unit efficiently.

Our objective is to implement the architecture proposed by Lang/Bruguera but with little change to facilitate the implementation. Base on the research on traditional double precision floating-point FMA arithmetic algorithms, a new high performance QPFMA unit is designed and realized, which supports four types of floating-point FMA operations 1 (multiply-add, multiply-sub, negative multiply-add and negative multiply-sub) along with floating-point multiplication, addition and comparison operation.

1. IEEE FLOATING POINT REPRESENTATION :

IEEE 754 is an IEEE Standard for Floating-Point Arithmetic which is developed for binary floating point arithmetic in 1985. In 1987 a second complementary standard (IEEE 854) is developed for radix independent floating point arithmetic[4]. Table 1 shows the standard five basic formats, named as single, double and quadruple precision basic formats.

Table 2-1: Basic formats of IEEE standard

Precision	No.of bits	Exponent	Mantissa + Hidden bit	Emin	Emax
Single	Binary 32	8	24	-126	+127
Double	Binary 64	11	53	-1022	+1023
Quadruple	Binary 128	15	113	-16382	+16383

Quadruple Precision floating-point data format:

IEEE 754-2008 floating point standard [1] defines 128-bit QP floating-point data type (binary128), as shown in Figure1, it consisting of three fields as:

- Sign (S) 1-bit,
- Exponent (E) 15-bit
- Fraction13 bits (112 explicitly stored)

The bias of exponent is 16383. The fraction of normalized data implies the integer bit as 1, which is not need to present.

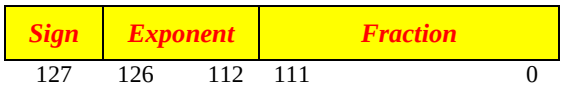

Fig1: Quadruple Precision floating-point data format

2. RELATED RESEARCH ON QP:

Floating point arithmetic is widely used in many areas. IEEE Standard 754 floating point is the most common repre-sentation today for real numbers on computers. The double precision floating point mul-tiplier using VHDL was synthesis in paper[1] with pipelining technique for Synthesis.

The architecture proposed by Lang/Bruguera was implement in paper[2], with the basic algorithm in the Verilog hardware description language, and then were synthesized, placed and routed for Cyclone II FPGA device using Quartus II 9.1. Area and timing information for each design approach was reported and analyzed.

[3]This paper also shows the full design of some blocks which are not included in Lang/Bruguera work like sign detection module.

This algorithm and a high performance QPFMA was designed and realized, which supports multiple floating-point arithmetic with a 7 cycles pipeline. By adopting dual adder and improving on algorithm architecture, optimizing leading zero anticipation and normalization shifter logic, the latency and hardware cost was decreased. Based on 65nm technology, the synthesis results show that the QPFMA can work at 1.2GHz, with the latency decreased by 3 cycles, the gate number reduced by 18.77% and the frequency increased about 11.63% comparing to current

QPFMA design, satisfying the requirements of high performance processor design.

[4]This paper presents the first hardware implementation of a fully parallel decimal floating-point fused-multiply-add unit performing the operation ± (A × B) ± C on decimal floating-point operands. The proposed design was fully compliant with the IEEE 754-2008 standard and supports the two standard formats decimal64 and decimal128.

In brief, QP floating-point arithmetic is one of the important trends of floating-point unit, helping to increase the precision of floating-point arithmetic and improve the performance of some important scientific applications. QP floating-point arithmetic put emphasis on theory exploration, rather from industry implementation. Considering of hardware and latency, the implementation QP floating-point arithmetic is not attractive for big area and long latency.

3. Standard Floating Point Fused Multiply-Add Algorithm:

A fused multiply add unit performs the multiplication B×C followed immediately by an addition of product and a third operand A so that the calculation of (A+ B × C) is done as a single and indivisible operation. It is also capable of performing ± (A± B × C) which supports four types of floating-point FMA operations 1 (multiply-add, multiply-sub, negative multiply-add and negative multiply-sub) along with floating-point multiplication, addition and comparison operation.

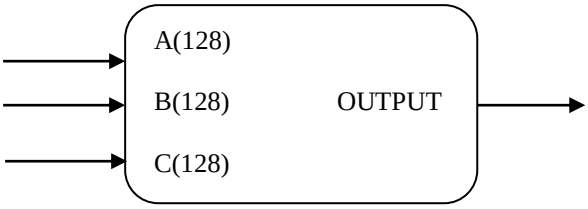

Fig2: A fused multiply add unit

3.1 Basic Algorithm :

LetA,B and C be the operands represented by (Ma ,Ea), (Mb ,Eb) and (Mc ,Ec) respectively. The significand are signed and normalized, and the OUT PUT result O is given by:

1. Multiply mantissa *Mb*and (Md= Mb X MC), add exponents *E*b and Ec (Ed=Eb+Ec)

2. Compare exponent Ea and Ed .Determine the shift of mantissa Ma.

3. Add/Sub the product Md and the aligned Ma.

4. Give the output of adder to Leading Zero Anticipator and Normalize the adder output and update the result exponent.

5. Round.

3.2 Basic implementation :

Using the above algorithm, the standard floating point fused multiply-add is designed. The organization of a FMA unit is shown in Figure (3). We can see that total cycles used are 5. The standard architecture is the baseline algorithm for floating-point fused multiply-add in any kind of hardware and software design.

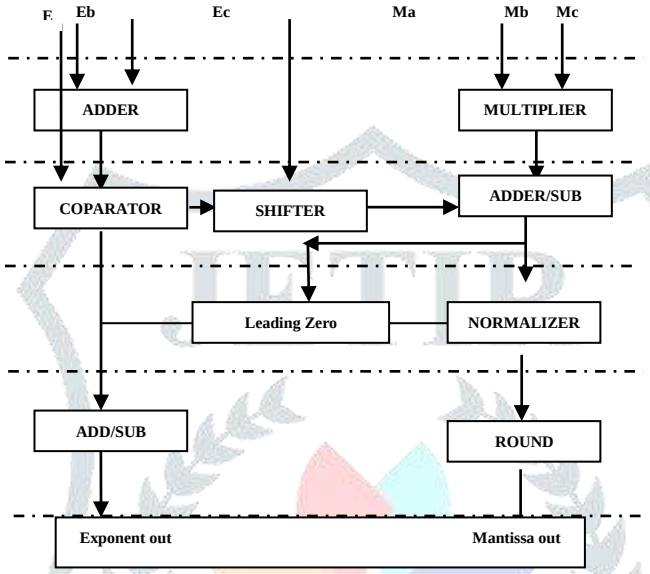

Fig3: Proposed Quadruple FMA unit

Description:

- **Cycle1:** Here we are doing addition of exponent of B and C , and simultaneously we are multiplying the mantissa of B and C.
- **Cycle2:** Here we are comparing the exponent of Ea with the Ed (Eb+Ec). Also we are shifting the mantissa of A according to the exponent adjustments and the we are doning Addition / Subtraction of the mantissas.
- **Cycle3:** Here the result of adder/sub is given for LZA and finally for normalization. The output of LZA is also given to exponent adjustment block.
- **Cycle4:** After rounding up mantissa and exponent we get final output .

4. CONCLUSION

The aim of this paper is to implement quadraple-precision binary floating point fused multiply add unit so here we give an overview of IEEE standard of floating point format with focusing on quadraple-precision binary floating point format. We also briefly explain the standard algorithm of the fused multiply add operation. Here we try to implement the FMA unit within 4 cycles.

REFERENCES :

[1] Sukhvir Kaur, Parminder Singh Jassal "Synthesis Of Double Precision Float-Ing Point Multiplier Using VHDL ", International Society of Thesis Publication Journal of Research in Electrical and Electronics Engineering (ISTP-JREEE).
[2] Eng. Walaa Abd El Aziz Ibrahim "Binary Floating Point Fused Multiply Add Unit", 3 Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013) .
[3] Rodina Samy, Hossam A. H. Fahmy, Ramy Raafat, Amira Mohamed, Tarek ElDeeb, Yasmin Farouk SilMinds, LLC. Maadi, "A Decimal Floating-Point Fused-Multiply-Add Unit" Electronics and Communication Department, Cairo University, Egypt.

[4] Mamidi Nagaraju,Geedimatla Shekar, "FPGA Based Quadruple Precision Floating Point Arithmetic for Scientific Computations International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-2 Number-3 Issue-5 September-2012